AD 630 471

STANDARDIZATION OF TYPEWRITER FONTS FOR AUTOMATIC READING

James F. Greenly

Rome Air Development Center
Griffiss Air Force Base, New York

March 1966

STANDARDIZATION OF TYPEWRITER FONTS FOR AUTOMATIC READING

James F. Greenly

Distribution of this document is unlimited

AFLC, GAFB, N.Y., 5 Apr 66-66

FOREWORD

This is a report of investigation and evaluation of existing typewriter fonts and related features (ribbons, papers, inks) which directly affect utilization of optical character recognition (OCR) equipments.

Engineering Practice Study Project No. MISC-0230 was established by Defense Supply Agency as a result of request from Defense Communications Agency to standardize typewriter fonts within the Department of Defense with regard to automatic reading requirements.

Partial funding for the project was established by Rome Air Development Center under Discretionary Funds Project DW-63-48, entitled "Standardization of Typewriter Fonts for Automatic Reading."

The major part of the testing and evaluation was performed by Link Group, Systems Division, General Precision Incorporated, Binghamton, N. Y., under Air Force Contract AF30(602)-3116.

This technical report covers research performed from February 1963 to September 1965.

RADC Project Engineer is James F. Greenly (EMIIH).

This technical report has been reviewed and is approved.

Approved: Frank J. Tomaini
FRANK J. TOMAINI
Chief, Info Processing Branch
Intel and Info Processing Div.

Approved: fr Howard Davis
ROBERT J. QUINN, JR., Colonel, USAF
Chief, Intel and Info Processing Div.

FOR THE COMMANDER: Irving J. Gabelman
IRVING J. GABELMAN
Chief, Advanced Studies Group

ABSTRACT

The objective of this program was to provide engineering data in support of standardization of typewriter fonts and related features for optical scanning application. Primary emphasis was placed on investigation and evaluation of existing typewriter fonts and includes an evaluation of a type font developed by Subcommittee X3.1 on Character Recognition under American Standards Association Sectional Committee X3.[1] Investigations were by computer programmed assessment of each font using a technique developed partly under Contract AF 30(602)-2642 sponsored by Rome Air Development Center and partly under continuing Link sponsored character recognition efforts. Evaluations were accomplished by extending the vocabulary capacity of a Link Multifont Page Reader to permit machine reading of a significant volume of typewriter-prepared documents. Reject and error rates were determined in this manner for each of several type styles considered.

[1] The American Standards Association does not officially endorse this font as a standard. In fact, when such an endorsement is made, the font will most probably be changed in many respects.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

SECTION I

INTRODUCTION

The following discussion describes the total effort in the area of Optical Character Recognition (OCR) standardization. It should be pointed out that, while the interests of the American Standards Association and the International Standards Organization include all types of printed characters, Rome Air Development Center confines its area of standardization to typewritten characters only.

1. AMERICAN STANDARDS ASSOCIATION

Sectional Committee X3 (Data Processing) formed Subcommittee X3.1 (Optical Character Recognition) in 1961 to promulgate standards for character sets and related features as applied to optical character recognition. Rome Air Development Center has held active membership in the task group associated with font specification. This group has done a rather thorough job in specifying the important parameters. However, because of the specialized commercial interests that are represented, the character sets that have evolved are quite highly stylized, and are favored more by the machine than by the human. Nevertheless, the character sets that have resulted represent a usable compromise for applications where accuracy and economy are of more importance than human readability.

To date, the group has finalized a set of characters consisting of numerals and a few symbols. A second set of characters, consisting of alphabetic upper case, numerals and symbols of set 1, and punctuation, is still in the process of finalization. In addition to the character shapes, Committee X3 has proposed standards for paper, print quality, and document format. There is very little interest in developing a standard for lower case characters because of limited commercial application.

2. EUROPEAN COMPUTER MANUFACTURERS ASSOCIATION (ECMA) AND THE INTERNATIONAL STANDARDS ORGANIZATION (ISO)

The ECMA was organized in 1961 in Geneva for the main purpose of promoting standards for data processing systems. Technical Committee TC4 was formed within the framework to standardize characters for optical recognition. The ECMA represents the majority group of computer users and manufacturers in Europe and, as a liaison member, works closely with the International Standards Organization in Geneva. Resulting standards are submitted as draft proposals to the ISO.

Liaison has been established by ECMA and ISO TC97 with ASA Committee X3. After a few joint meetings during 1963 and 1964, reasonably close agreement was reached concerning a numerals character set, designed and proposed by the ASA group. To date, agreement has not been reached concerning an alphabetic character set.

3. ROME AIR DEVELOPMENT CENTER EFFORT

In 1962, the Department of Defense recognized a future requirement for standardizing typewriter fonts in order to facilitate use of optical character recognition

devices, and initiated a project, naming Rome Air Development Center as the action agency to standardize typewriter fonts within the DOD. Accordingly, RADC contracted with Link Division, G.P.I., to screen the available typewriter fonts, and to perform actual machine reading tests on a group of candidate fonts.

RADC has established a background, maintaining a continuing program in the area of optical character recognition since 1957, placing primary emphasis in the development of page reading devices for both Russian and English.

SECTION II

THE LINK READING TECHNIQUE

Link Division, G.P.I., for several years, has been actively engaged in programs directed towards the development of advanced optical character recognition techniques. Systems have been developed with which successful reading of a variety of type styles has been demonstrated. Several reading techniques have been investigated; however, the most successful one uses an electronic peephole matching principal in which only selected subareas of the image field are used as apertures for matching. The method allows complete disregard for serifs and other typographic and stylistic embellishments, resulting in a simple and reliable system for multifont reading.

Each character is scanned by a column of photodetectors and converted into a digital waveform from which selected portions are matched against a reference vocabulary, the recognition criteria being the least value of the time integrals of the total number of absolute differences between the incoming video and stored descriptions of each character in the vocabulary.

1. THE LINK MULTIFONT PAGE READER

Link has developed a multifont optical page reading system for commercial applications which can convert pages of printed or typewritten text into computer language. The device features automatic page transporting, line location, and scanning.

In a given OCR application, the degree of control to be exercised over input documents is a very basic decision. Many times it is possible to specify the type style(s) permissible for the preparation of these documents. In each of the general categories of unstylized, semi-stylized, and stylized fonts a recommended font is presented, its selection justified, and its performance demonstrated using an existing optical page reader.

A Link Multifont Page Reader is used as a test vehicle for demonstrating the results of the evaluation. For evaluating reading machine performance printout, an output typewriter is provided.

As a part of this program, the vocabulary of the system was modified to include three fonts: unstylized Artisan 10, semi-stylized Manifold 12, and a stylized type font developed by Subcommittee X3.1 on Character Recognition under American Standards Association Sectional Committee X3. The Link Page Reader, however, did not dictate the selection of these particular fonts, since the selection criteria used were applicable to all reading techniques based on area analysis. Two additional type fonts, Boldface #16 and Financial Gothic, were investigated without hardware evaluation.

The existing vocabulary of this system was extended to accommodate the additional fonts that were required by RADC. The equipment stores an entire page in a magnetic core memory, and then prints out into an electric typewriter for verification.

2. MODIFICATIONS MADE FOR TYPE FONT EVALUATIONS

The addition of two type fonts, i.e., Artisan 10 and ASA Candidate Font to the Link Page Reader required minor modifications to the Recognition Unit and to the Data Pickup Unit.

The decision making circuits, (called "Output Selection Latches") were expanded in total number to accommodate the additional fonts to be evaluated. The system had previously been designed to read two type fonts one of which consisted of 55 characters while the second included 19 characters. The memory for each of these fonts and for each of the fonts investigated under the subject program is operator selectable.

A change in optical system magnification was found necessary in order to meet resolution requirements and at the same time provide reasonable registration tolerance when reading the fonts added. Scanning resolution was determined by stroke width. A vertical registration shift tolerance of ±1 scanning raster row is obtained by adjusting the magnification of the optical system to the point where at least three rows of the scanning raster represent a horizontal stroke thickness. Using only bits from the center row then, a ±1 row shift in the vertical registration can be tolerated without loss of recognition accuracy. System magnification then, is fixed by the thickness of a horizontal stroke. Maximum height of characters ranged between .100 and .130 inches with the associated scanning raster consisting of 20 to 30 rows of video, respectively. For the larger size, i.e. the ASA Candidate Font, a magnification of 17.6 is used while magnification for reading Artisan 10 or Manifold 12 is set at 20.4. Change in magnification requires a small change in illumination level to cause the photodiode outputs to stay constant for a given range of stroke density.

Data for encoding Bold Face #16, Underwood Financial Gothic and Artisan 10 was taken on an earlier Link Page Reader known as the Model X-3. This system is limited to a lower maximum allowable resolution than the present multifont page reader, i.e. 250 as compared to 480 areas per character matrix. While data was processed for certain of the above mentioned type styles, using the Model X-3 system, development of an improved Link Page Reader progressed to a state where certain logic refinements in this system made it desirable to perform further evaluations using this improved equipment. Data was therefore, processed using this newly developed machine for the two type styles, selected for readability demonstration. A distinction must be made therefore, between the low resolution data of Artisan 10 (1) and the higher resolution data of Artisan 10 (2), the former taken on the Model X-3 and the latter on the newly developed page reader.

SECTION III

THE FONT ASSESSMENT TECHNIQUE

1. AUTOMATIC CHARACTER ANALYZER

The reliability of any reading technique based on comparisons of characteristics of unknown patterns with stored information must depend to a large extent on the accuracy of the stored information. To the highest degree possible, this information must include only nonvariant characteristics of the specific patterns to be recognized, disregarding all variables introduced by machine peculiarities, ribbon condition, or human fallibility.

One of the early methods for solving this problem in the Link development program was to project a magnified character image onto an especially designed overlay representing the reader scanning format. Each incremental area (bit) of the character was then assigned a corresponding weight based on the relative amount of black or white information contained in the area in question and the surrounding areas. Several weaknesses existed in this technique, however, including the following:

1) In order to obtain a usable image size, a transparency was made of each character, permitting magnification by projection. The integration of character detail in photographic processes introduced error.

2) The establishment of scanning references and bit color was a matter of human judgment and was, therefore, subject to error.

3) The data had to be based on only a few impressions of each character because of the laborious and time-consuming work involved. Therefore, differences due to type variations, ribbon condition, etc., arising in large data samples were not fully accounted for.

4) No accurate method of simulating possible worst-case conditions existed.

Additional development, however, resulted in the automatic system for data taking, known as the Automatic Character Analyzer.

The Analyzer utilizes the image transducer and timing signals from the Reader electronics to accumulate statistical data; specifically, the probability of black and/or white occurrence in every incremental area (bit) and many samples of a particular typed or printed character.

Scanning references and bit color are no longer subject to human decisions, since they are established by the reading machine. A quick and accurate composite from many samples of each character from a variety of machines of the type styles under consideration can now be made. In multifont applications, determination of the amount of invariant data for the same character in two or more type styles can be done by preparing documents having equal numbers of each type sample. These are rapidly converted into the desired composite for samples of 20, 100, or 1,000 characters.

In addition to its value in determining memory encoding information, the Analyzer is a powerful development tool. It provides immediate research data on the value of various reader logic design innovations.

The Automatic Character Analyzer operation results in the following outputs:

1) Counting and displaying the total occurrence of white for each bit position for a character sample of 20, 100, or 1,000 impressions.

2) Counting and displaying the total occurrence of black for each bit position for a character sample of 20, 100, or 1,000 impressions.

3) Selecting and displaying the appropriate row (1 to 30) and column (A to P) locations being analyzed at any particular time.

4) Providing a sequential printed output, including column and row identification, total white occurrence, and total black occurrence for each bit position of the scanning format.

The next logical improvement for this system would be a conversion unit for production of punched card outputs. The cards, punched in appropriate coded form, could be immediately processed by the IBM 1460 computer facility at Link for the purpose of reader memory encoding.

SECTION IV

FONT INVESTIGATIONS

Performance figures (Figure 9) have been obtained using a Link Page Reader equipped for type font data reduction and modified to include the additional fonts of Figure 10 in its vocabulary for reading performance evaluation. In examining the performance figures shown in Figure 9, certain factors should be kept in mind when comparing reject and error rates for unstylized, semi-stylized, and stylized vocabularies. Test documents for the ASA Candidate Font and for Artisan 10 were limited by economic considerations to single typewriters for their preparation while approximately 30 typewriters and as many different operators were involved in the preparation of Modified Manifold 12 samples.

The absence of multiple typewriter and multiple operator-induced variables in the original type samples on which the reading machine memory is based, and in the test documents used for determining reject and error rates, undoubtedly makes the Artisan 10 and ASA Font performance data appear better than would occur in a practical application. Modified Manifold 12 results then are probably most representative of performance achievable under "real world" handicaps. A common basis exists, however, between the ASA Candidate font and Artisan 10 results, which permit a direct comparison and a good indication of the accuracy improvement to be gained by stylization. Again, however, this conclusion must be modified by the larger number of characters in the unstylized font tested since, as the number of characters to be recognized is increased, the probability that two or more characters will look alike also increases, making recognition more difficult.

It was interesting to note during the preparation and testing of these documents that the number of errors made by trained typists far exceeded the number of reading machine errors made as a result of reading these same documents.

Certain system features were disabled during this program in order to prevent influencing results with an unnecessary number of variables. These features were:

1) Automatic paper feed, loading, discharge or stacking.
2) Automatic line finding or tracking (line skew compensation).
3) Line rescans in an attempt to recognize rejects.
4) Magnetic tape output unit.

All test documents (see actual type samples with associated printouts shown in Figure 11) were prepared on one grade of paper using polyethylene carbon ribbon equipped electric typewriters. The paper used was one selected as being most suitable for optical character recognition applications (see Section V.1.d.). The reflectivity of this paper is high, and the adherence of ink to the paper when using polyethylene carbon ribbons is uniform. No typewriter cleaning or other typewriter maintenance procedures were carried out prior to or during the reading performance evaluation.

Documents are mounted manually on the scanning mechanism and positioning of the optical system successively over each line to be read is also operator controlled.

970	820	18	25	51			
2 287	587	16	25	51			
2 302	091	17	24	24.1	25	55	57.1
2 322	435	15	22.2	25	51	54	
2 451	465	15	25	33.1	33.2	56	59
2 501	788	10	15	16	21.3	25	34.3
2 540	660	15	25	34.3	51	54	
2 628	689	14	16	21.1	21.6	22.2	25
2 645	741	16	18	25	56		
2 646	465	15	25	31.2	33.1	34.3	41.1
2 664	243	16	21.4	25	33.2	56	59
2 677	728	15	23.1	25	32	51	54
2 685	615	14	15	21.3	21.6	25	34.2
2 693	908	16	18	21.3	21.6	23.1	25
2 779	428	16	21.3	23.3	23.2	15	33.2
2 800	649	16	21.6	25	28	33.8	56
2 855	147	16	25	33.2	33.8	56	59
2 872	996	16	13.1	25	33	33.5	33.8
2 878	999	16	25	33	33.2	56	59
2 907	400	61	25	33.2	33.8	56	
2 908	761	15	25	33.2	34.3	51	56
2 908	889	16	13.1	13.1	33	33.2	56
2 912	164	16	23.1	23	33.8	56	
2 921	738	16	25	33	33.2	56	
2 927	656	16	23.2	25	33	33.2	33.5
2 983	900	16	18	21.4	25	34.3	51
2 986	606	16	25	51	56		
2 896	722	16	23.1	25	28	51	56
2 989	726	16	25	33.2	56		
2 991	446	16	33.2	25	33.2	56	
2 999	196	18	25	35	52		

Fig.1 American Standards Association Numeric Font TG1

970 820	18	25	51			
2 287 587	16	25	51			
2 302 081	17	24	24.1	25	55	57.1
2 322 435	15	22.2	25	51	54	
2 451 465	15	25	33.1	33.2	56	59
2 501 788	10	15	16	21.3	25	34.3
2 540 660	15	25	34.3	51	54	
2 628 689	14	16	21.1	21.6	22.2	25
2 645 741	16	18	25	56		
2 646 465	15	25	31.2	33.1	34.3	41.1
2 664 243	16	21.4	25	33.2	56	59
2 677 728	15	23.1	25	32	51	54
2 685 615	14	15	21.3	21.6	25	34.2
2 693 908	16	18	21.3	21.6	23.1	25
2 779 428	16	21.3	23.2	25	33.2	35
2 800 639	16	21.6	25	28	33.8	56
2 855 147	16	25	33.2	33.8	56	59
2 872 996	16	13.1	25	33	33.5	33.8
2 878 999	16	25	33	33.2	56	59
2 907 400	61	25	33.2	33.8	56	
2 908 761	15	25	33.2	34.3	51	56
2 908 889	16	13.1	25	33	33.2	56
2 912 164	16	23.1	25	33.8	56	
2 921 738	16	25	33	33.2	56	
2 927 656	16	23.2	25	33	33.2	33.5
2 983 900	16	18	21.4	25	34.3	51
2 986 606	16	25	51	56		
2 896 722	16	23.1	25	28	51	56
2 989 726	16	25	33.2	56		
2 991 447	16	33.2	25	33.2	56	
2 999 196	18	25	35	52		

Fig. 2 I.B.M. 1428 Numeric Font

2	970	820	18	25	51				
2	287	578	16	25	51				
2	302	081	17	24	24.1	25	55	57.1	
2	322	435	15	22.2	25	51	54		
2	451	465	15	25	33.1	33.2	56	59	
2	501	788	10	15	16	21.3	25	34.3	
2	540	660	15	25	34.3	51	54		
2	628	689	14	16	21.1	21.6	22.1	25	
2	645	741	16	18	25	56			
2	646	465	15	25	31.2	33.1	34.3	41.1	
2	664	243	16	21.4	25	33.2	56	59	
2	677	728	15	23.1	25	32	51	54	
2	685	615	14	15	21.3	21.6	25	34.2	
2	693	908	16	18	21.3	21.6	23.1	25	
2	779	428	16	21.3	23.2	25	33.2	35	
2	800	639	16	21.6	25	28	33.8	56	
2	855	147	16	25	33.2	33.8	56	59	
2	872	996	16	23.1	25	33	33.5	33.8	
2	878	999	16	55	33	33.2	56	59	
2	907	400	16	25	33.3	33.8	56		
2	908	761	15	25	25	33.2	34.3	51	
2	908	899	16	13.1	25	33	33.2	56	
2	912	164	16	23.1	25	33.8	56		
2	921	738	16	25	33	33.2	56		
2	927	656	16	23.1	25	33	33.2	33.5	
2	983	900	16	18	21.4	25	34.3	51	
2	986	606	16	25	51	56			
2	896	722	16	23.1	23.1	25	28	51	
2	989	726	16	25	33.2	56			
2	991	447	16	33.2	25	33.2	56		
2	999	196	18	25	35	52			

Fig.3 Farrington Selfchek

2 372 882	PHILLIPS	45	IBM	CARD SENSING MACHINE
2 779 811	PICIANO	57	VITRO	PHOTOCELL CONSTRUCTION
2 908 889	PIETY	59	PHILLIP	COMPUTER
2 915 246	PIETY	59	PHILLIP	POLYNOMIAL ROOTS COMPUTER
2 615 992	PIKE	52	RCA	APPARATUS FOR INDICIA RECOGNITION
859 117	POLLARD	61	FERRANTI	GREAT BRITAIN
2 432 123	POTTER	47	BTL	TRANSLATION VISUAL SYMBOLS
2 692 551	POTTER	54		HIGH SPEED ROTARY PRINTER
553 500	POTTER	57	WE	BELGUIM
2 177 017	POTTS	39	TT	PHOTOELECTRIC TRANSMITTER
2 274 737	POTTS	42	TT	PHOTOELECTRIC TRANSMITTER
2 360 589	POTTS	44	TT	TELEGRAPH PRINT APPARATUS
2 565 266	POTTS	51	TT	HIGH SPEED PHOTOELECTRIC TRANS
2 586 711	POTTS	52	TT	SCANNING SYSTEM APPARATUS
2 613 809	POTTS	52	TT	TICKET SORTING APPARATUS
1 746 331	PRASHKER	30		MECHANICAL TELAUTOGRAPH
1 215 000	QUADE	59	IBM	FRANCE
2 875 429	QUADE	59	IBM	PHASE SENSITIVE MAGNETIC HEAD
1 487 115	QUARRIE	24	WE	INTELLIGENCE SYSTEM
2 942 237	QUIOGUE	60	BURRU	SIGNAL GENERATOR CONTROL
2 929 047	RABINOW	60	SPERRY	RADAR TARGET CLASSIFICATION
2 933 246	RABINOW	60		READING MACHINE
2 795 705	RABINOW	57	COMMER	OPTICAL COINCEDENCE DEVICES
2 810 097	RABINOW	57		OPTIMUM SENSITIVITY AUTOMATIC
468 448	RADLEY	37		GREAT BRITAIN
2 908 761	RAISBECK	59	BTL	VOICE PITCH DETERMINATION
2 854 191	RAISBECK	58	BTL	COMPUTATION OF CORRELATION
2 667 599	RAJCHMAN	54	RCA	ELECTRONIC SWITCHING DEVICE
2 742 631	RAJCHMAN	56	RCA	RECORD + TRANSMIT PHOSPHORS
2 788 879	RAND	57	SPERRY	INFORMATION RECORDING APPARATUS
2 982 899	RAPPAPORT	61	SPERRY	PRESSURE SENSITIVE FOLLOWER
2 458 030	REA	49	BTL	SELECTIVE SIGNALING APPARATUS

Fig. 4 American Standards Association Alphabetic Font TG1C
Total Characters- ABCDEFGHIJKLMNOPQRSTUVWXYZ
⑆$⑇.,-?/ 1234567890

2	372 882	PHILLIPS	45	IBM	CARD SENSING MACHINE
2	779 811	PICIANO	57	VITRO	PHOTOCELL CONSTRUCTION
2	908 889	PIETY	59	PHILLIP	COMPUTER
2	915 246	PIETY	59	PHILLIP	POLYNOMIAL ROOTS COMPUTER
2	615 992	PIKW	52	RCA	APPARATUS FOR INDICIA REC
	859 117	POKLARD	61	FERRANTI	GREAT BRITAIN
2	432 123	POTTER	47	BTL	TRANSLATION VISUAL METHODS
2	692 551	POTTER	54		HIGH SPEED ROTARY PRINTER
	553 500	POTTER	57	WE	BELGIUM
2	177 017	POTS	39	TT	PHOTOELECTRIC TRANSMITTER
2	274 737	POTTS	42	TT	PHOTOELECTRIC TRANSMITTER
2	360 589	POTTS	44	TT	TELEGRAPH PRINT APPARATUS
2	565 266	POTTS	51	TT	HIGH SPEED PHOTOELECTRIC
2	586 711	POTTS	52	TT	SCANNING SYSTEM APPARATUS
2	613 809	POTTS	52	TT	TICKET SORTING APPARATUS
1	746 331	PRASHKER	30		MECHANICAL TELAUTOGRAPH
1	215 000	QWADE	59	IBM	FRANCE
2	875 429	QUADE	59	IBM	PHASE SENSITIVE MAGNETIC
1	487 115	QUARRIE	24	WE	INTELLIGENCE SYSTEM
2	942 237	QUIOGUE	60	BUURO	SIGNAL GENERATOR CONTROL
2	929 047	RABINOW	60	SPERRY	RADAR TARGET CLASSIFICATION
2	933 246	RABINOW	60		READING MACHINE
2	795 705	RABINOW	57	COMMER	OPTIVAL COINCIDENCE DEVICE
2	810 097	RABINOW	57		OPTIMUM SENSITUVITY AUTOM
	468 448	RADLEY	37		GREAT BRITAIN
2	908 761	RAISBECK	59	BTL	VOICE PITCH DETERMINATION
2	854 191	RAISBECK	58	BTL	COMPUTATION OF CORRELATION
2	667 599	RAJCHMAN	54	RCA	ELECTRONIC SWITCHING DEVICE
2	742 631	RAJCHMAN	56	RCA	RECORD-TRANSMIT PHOSPHORS
2	788 879	RAND	57	SPERRY	INFORMATION RECORDING DEV
2	982 899	RAPPAPORT	61	SPERRY	PRESSURE SENSITIVE FOLLOWER
2	458 030	REA	49	BTL	SELECTIVE SIGNALING DEVICE

Fig.5 I.B.M. 1428 Alpha-numeric Font
TOTAL CHARACTERS- ABCDEFGHIJKLMNOPQRSTUVWXYZ
1234567890-=×+$,./

2 372 882	PHILLIPS	45	IBM	CARD SENSING MACHINE
2 779 811	PICIANO	57	VITRO	PHOTOCELL CONSTRUCTION
2 908 889	PIETY	59	PHILLIP	COMPUTER
2 915 246	PIETY	59	PHILLIP	POLYNOMIAL ROOTS COMPUTER
2 615 992	PIKE	52	RCA	APPARATUS FOR RECOGNITION
859 117	POLLARD	61	FERRANTI	GREAT BRITAIN
2 432 123	POTTER	47	BTL	TRANSLATION VISUAL SYMBOLS
2 692 551	POTTER	54		HIGH SPEED ROTARY PRINTER
553 500	POTTER	57	WE	BELGUIM
2 177 017	POTTS	39	TT	PHOTOELECTRIC TRANSMITTER
2 274 737	POTTS	42	TT	PHOTOELECTRIC TRANSMITTER
2 360 589	POTTS	44	TT	TELEGRAPH PRINT APPARATUS
2 565 266	POTTS	51	TT	HIGH SPEED PHOTOELEC TRANS
2 586 711	POTTS	52	TT	SCANNING SYSTEM APPARATUS
2 613 809	POTTS	52	TT	TICKET SORTING APPARATUS
1 746 331	PRASHKER	30		MECHANICAL TELAUTOGRAPH
1 215 000	QUADE	59	IBM	FRANCE
2 875 429	QUADE	59	IBM	PHASE SENSITIVE MAG HEAD
1 487 115	QUARRIE	24	WE	INTELLIGENCE SYSTEM
2 942 237	QUIOGUE	60	BURRO	SIGNAL GENERATOR CONTROL
2 929 047	RABINOW	60	SPERRY	RADAR TARGET CLASSIFICATION
2 933 246	RABINOW	66		READING MACHINE
2 795 705	RABINOW	57	COMMER	OPTICAL COINCIDENCE DEVICE
2 810 097	RABINOW	57		OPTIMUM SENSITIVITY AUTOM
468 448	RADLEY	37		GREAT BRITAIN
2 908 761	RAISBECK	59	BTL	VOICE PITCH DETERMINATION
2 854 191	RAISBECK	58	BTL	COMPUTATION OF CORRELATION
2 667 599	RAJCHMAN	54	RCA	ELECTRONIC SWITCHING DEV
2 742 631	RAJCHMAN	56	RCA	RECORD + TRANSMIT PHOSPHOR
2 788 879	RAND	57	SPERRY	INFORMATION RECORDING
2 982 899	RAPPAPORT	61	SPERRY	PRESSURE SENSITIVE FOLLOWER
2 458 030	RCA	49	BTL	SELECTIVE SIGNALING

Fig. 6 FARRINGTON ALPHA-NUMERIC FONT 12L
Total Characters-ABCDEFGHIJKLMNOPQRSTUVWXYZ
1234567890/().,#'$

2	970	820	18	25	51			
2	287	578	16	25	51			
2	302	081	17	24	24.1	25	55	57.1
2	322	435	15	22.2	25	51	54	
2	451	465	15	25	33.1	33.2	56	59
2	501	788	10	15	16	21.3	25	34.3
2	540	660	15	25	34.3	51	54	
2	628	689	14	16	21.1	21.6	22.1	25
2	645	741	16	18	25	56		
2	646	465	15	25	31.2	33.1	34.3	41.1
2	664	243	16	21.4	25	33.2	56	59
2	677	728	15	23.1	25	32	51	54
2	685	615	14	15	21.3	21.6	25	34.2
2	693	908	16	18	21.3	21.6	23.1	25
2	779	428	16	21.3	23.2	25	33.2	35
2	800	639	16	21.6	25	28	33.8	56
2	855	147	16	25	33.2	33.8	56	59
2	872	996	16	23.1	25	33	33.5	33.8
2	878	999	16	55	33	33.2	56	59
2	907	400	16	25	33.3	33.8	56	
2	908	761	15	25	25	33.2	34.3	51
2	908	899	16	13.1	25	33	33.2	56
2	912	164	16	23.1	25	33.8	56	
2	921	738	16	25	33	33.2	56	
2	927	656	16	23.1	25	33	33.2	33.5
2	983	900	16	18	21.4	25	34.3	51
2	986	606	16	25	51	56		
2	896	722	16	23.1	23.1	25	28	51
2	989	726	16	25	33.2	56		
2	991	447	16	33.2	25	33.2	56	
2	999	196	18	25	35	52		

Fig. 7 I.B.M. Pica Numeric Font

2	372	882	PHILLIPS	45	IBM	CARD SENSING MACHINE
2	779	811	PICIANO	57	VITRO	PHOTOCELL CONSTRUCTION
2	908	889	PIETY	59	PHILLIP	COMPUTER
2	915	246	PIETY	59	PHILLIP	POLYNOMIAL ROOTS COMPUTER
2	615	992	PIKE	52	RCA	APPARATUS FOR RECOGNITION
	859	117	POLLARD	61	FERRANTI	GREAT BRITAIN
2	432	123	POTTER	47	BTL	TRANSLATION VISUAL SYMBOLS
2	692	551	POTTER	54		HIGH SPEED ROTARY PRINTER
2	177	017	POTTS	39	TT	PHOTOELECTRIC TRANSMITTER
	553	500	POTTER	57	WE	BELGIUM
2	274	737	POTTS	42	TT	PHOTOELECTRIC TRANSMITTER
2	565	266	POTTS	51	TT	HIGH SPEED PHOTOELEC TRANS
2	586	711	POTTS	52	TT	SCANNING SYSTEM APPARATUS
2	613	809	POTTS	52	TT	TICKET SORTING APPARATUS
1	746	331	PRASHKER	30		MECHANICAL TELAUTOGRAPH
1	215	000	QUADE	59	IBM	FRANCE
2	875	429	QUADE	59	IBM	PHASE SENSITIVE MAG HEAD
1	487	115	QUARRIE	24	WE	INTELLIGENCE SYSTEM
2	942	237	QUIGUE	60	BURRO	SIGNAL GENERATOR CONTROL
2	929	047	RABINOW	60	SPERRY	RADAR TARGET CLASSIFICATION
2	933	246	RABINOW	66		READING MACHINE
2	795	705	RABINOW	57	COMMER	OPTICAL COINCIDENCE DEVICE
2	810	097	RABINOW	57		OPTIMUM SENSITIVITY AUTOM
	468	448	RADLEY	37		GREAT BRITAIN
2	908	761	RAISBECK	59	BTL	VOICE PITCH DETERMINATION
2	854	191	RAISBECK	58	BTL	COMPUTATION OF CORRELATION
2	667	599	RAJCHMAN	54	RCA	ELECTRONIC SWITCHING DEV
2	742	631	RAJCHMAN	56	RCA	RECORD + TRANSMIT PHOSPHOR
2	788	874	RAND	57	SPERRY	INFORMATION RECORDING
2	982	899	RAPPAPORT	61	SPERRY	PRESSURE SENSITIVE FOLLOWER
2	458	030	RCA	49	BTL	SELECTIVE SIGNALING

Fig. 8 Artisan 10 Alpha-numeric Font (I.B.M. Numerals)

Figure 10. Comparative Samples of the Candidate Fonts Selected for Assessment

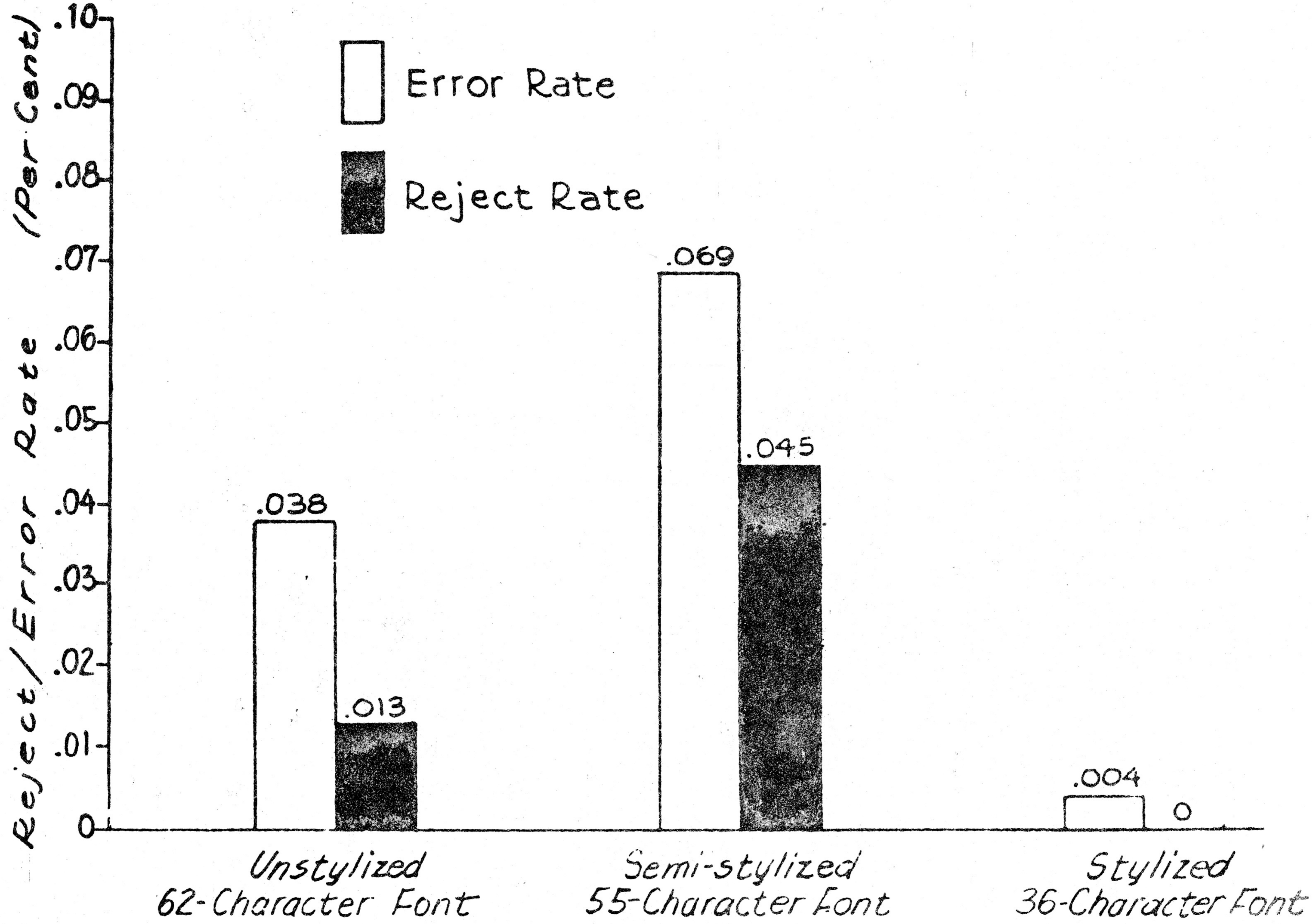

Figure 9. LINK Page Reader Performance While Reading Unstylized, Semi-Stylized and Stylized Type Fonts

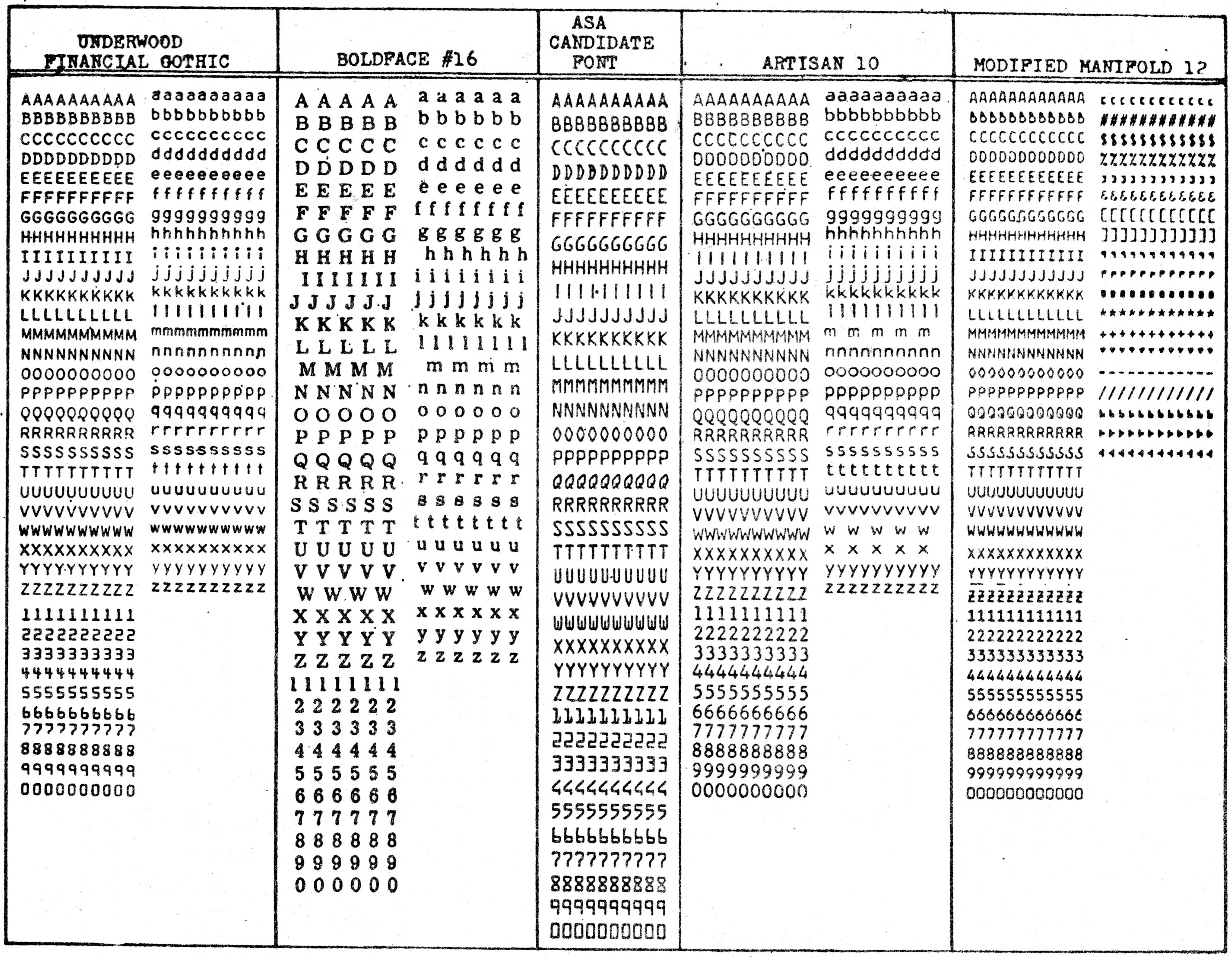

Figure 10. Comparative Samples of the Candidate Fonts Selected for Assessment

ARTISAN 10 ORIGINAL COPY

THE LINK READING TECHNIQUE

While performing the various character recognition programs, Link has become well aware of the attributes of characters produced in varying type styles and by a variety of printing devices. In the design of a reading machine, a multitude of variables must be considered, such as horizontal and verti cal registration, character skew, line skew, character reflec tance, document reflectance, character height, character width, stroke width,additive noise, reductive noise, voids, embossing, type face wear and type face manufacturing tolerances.

LINK PAGE READER PRINTOUT

```
THE LINK READING        TECHNIQUE ;
While performing the various character recognition programs, ;
Link has become well aware of the attributes of characters ;
produced in varying type styles and by a variety of printing ;
devices  In the design of a reading       machine, a multitude ;
of variables must be considered, such as horizontal and verti ;
cal registration, character skew, line skew, character reflec ;
tance, document reflectance, character height, character width, ;
stroke width,additive noise, reductive noise, voids, embossing, ;
type face wear and type face manufacturing tolerances
```

MODIFIED MANIFOLD 12 ORIGINAL COPY

82863D					
7000003	B/M ADV. TO BF	B/M		700 00	
7000003	B/M PMTS TO BF	B/M		558 48	
7000003	B/M CKS. TO BF	B/M		115 02	
2010007	P&L GREENWOOD #6633.30	P&L COLL	48 00		
33063D					
L	COLLECTIONS & LOANS		4 812 51	8 806 53	
9024802	UN.CHGS. IL LARGE [$COST]		1 469 15		
7270002	INS COST LIFE $81.79	INS	81 79		
C	EXCH.CK.#28 10076.8-LUSSIER		225 22	225 22	925

LINK PAGE READER PRINTOUT

```
82863d .          .           .         .            ;
7000003 .         b/m adv to bf b/m .   .            700 00 .      ;
7000003 .         b/m pmts to bf b/m .  .            558 48 .      ;
7000003 .         b/m cks to bf b/m .   .            115 02 .      ;
2010007 .         p&l greenwood #6633 30 p&l coll .  48 00 .       .            ;
83063d .          .           .         .            ;
1 .               collections & loans .  4 812 51 .  8 806 53 .    ;
8024002 .         un chgs il large ($cost) .         1 469 15. .   .            ;
7270002 .         ins cost life $81 79 ins .         81 79. .      .            ;
c .               exch ck #28 10076 8-lussier .      225 22 .      225 22.      925
```

ASA CANDIDATE FONT ORIGINAL COPY

11M AMER COC VOL 8 NO 3 PP 221 TO 226.

14M TO HELP THE INDIVIDUAL SCIENTISTS CONTROL THE LITERATURE

HE READS AND TO HELP THE SCIENTISTS KEEP INFORMED AS TO WHAT

OTHER WORKERS ARE DOING IN HIS FIELD RIGHT NOW THIS AURTHOR

SUGGESTS THAT THE SCIENTISTS MAINTAIN A READING LOG WHICH HE

LINK PAGE READER PRINTOUT

```
11m amer coc vol 8 no 3 pp 221 to 226 ;
;
14m to help the individual scientists control the literature ;
;
he reads and to help the scientists keep informed as to what ;
;
other workers are doing in his field right now this aurthor ;
;
suggests that the scientists maintain a reading log which he ;
```

Figure 11. Original Copy and Associated Printouts for Unsytlized, Semi-Stylized and Stylized Type Samples

Character scanning is at 250 characters per second. After each character is read, the machine decision is temporarily stored in a magnetic core buffer memory until the entire line has passed the photodetector array, followed by printout of the information on an input-output typewriter.

Each typewritten test document was scanned and printed out; then printout and original were visually compared with all rejects and errors noted. The reject and error rate goals (combined $\leq$ 0.2 percent) should, of course, apply only to those characters scanned which are machine chargeable rejects and errors. Otherwise, the figures do not provide an accurate indication of font or reading machine performance since they would be influenced by such things as typist proficiency, negligence of normal typewriter maintenance, mutilation of documents by improper manual handling, etc.

Each reject and error, however, has been considered Scanner chargeable except for flagrant violations on character quality. To be more specific, rejects and errors have been charged against the Scanner when caused by:

1) Poor erasures
2) Mild amounts of dirt
3) Small voids
4) Mild embossing
5) Low contrast
6) Wrinkles or creases
7) Unknown factors

Rejects or errors have not been charged against the Scanner when caused by:

1) Overstrikes or merge with adjacent characters
2) Very poor erasures
3) Severe amounts of dirt
4) Large voids
5) Holes or tears in reading area
6) Severe embossing
7) Severe line skew ($>$ ±30)
8) Operator carelessness when operating the Scanner.

1. SELECTION OF CANDIDATE FONTS FOR ASSESSMENT

From a survey of available typewriter fonts, a group of candidate fonts for automatic reading has been compiled as shown in Table I. Typewritten test samples of each were examined with certain characteristic measurements made for each. Based on the following selection criteria, the list of candidates was reduced to five fonts. From these, three were encoded (numerals and upper and lower case alphas) into the memory of a Link Page Reading System for reading performance evaluation using actual documents.

Candidate Fonts Selection Criteria

a) Characters within sets to have reasonably uniform overall height and width.

b) Characters within sets to have reasonably uniform stroke width.

c) Characters within sets to be reasonably pleasing in appearance.

d) The standard characters of each set should be given preference over options offered by typewriter manufacturers.

e) Only upper case characters are examined in detail; however, this did not limit selections to only single case fonts.

f) Sets in which merged and overlapping series occur frequently are avoided except where merge and overlap is consistently confined to specific locations (serifs).

g) Discrimination between "O" and "zero" and between "I" and "one" are not considered essential.

h) Fonts listed in the Federal Supply Catalog under Air Force stock class 7430 have been given serious consideration.

Using the above selection criteria, the type styles of Table I were reduced to the following five fonts, comparative samples of which are shown in Figure 10:

1. Boldface #16
2. Artisan 10
3. Underwood Financial Gothic
4. Modified Manifold 12
5. ASA Candidate Font for optical scanning application

In the selection of Artisan 10 as a candidate, the standard numerals of this font were judged to be undesirable for OCR because of height variations which increased the possibility of merge between adjacent lines. An optional numeral set, one considered standard for the IBM Prestige Pica font, is considered superior in the unstylized class for OCR; therefore, the test typewriter was equipped with Artisan 10 alphas and symbols with Prestige Pica numerals. This combination then, is the font subsequently referred to in this report as Artisan 10.

2. UNSTYLIZED FONT INVESTIGATIONS

Type fonts designed for commercial typewriters, with the principal design criterion of appearing pleasing to the human eye, are considered unstylized (i.e., not designed with mechanized optical character recognition considerations in mind). The great majority of type fonts fall into this category. The selection of such fonts as candidates for optical scanning must be based on characteristics that allow maximum recognition in a scanning system. Boldface #16 and Artisan 10 were selected as unstylized fonts well suited for optical scanning application.

TYPE STYLE	MANUFACTURER	PITCH	LINES/INCH	MEASURED STROKE WIDTH (Capital M)	MEASURED HEIGHT (Capital M)	MEASURED WIDTH (Capital M)	SPECIFIED CAPITAL HEIGHT (Die Dimension)	COMMENTS
[1]Elite 02	IBM	12	6	.013	.100	.075	.086	
[1]Standard Elite	Royal	12	6	.013	.103	.064		
[1]Executype 516*	Remington Rand	10	6	.010	.102	.065		
[1]Elite 203	Remington Rand	12	6	.010	.105	.068		
[1]Modern Gothic	IBM	9	5-1/4	.012	.127	.088	.114	
[1]Pica 400	Royal	10	6	.012	.110	.077		
[1]Pica 01	IBM	10	6	.013	.110	.080	.096	
[1]Pica Gothic	IBM	10	6	.012	.112	.087	.100	Sans Serif
[1]Pica Single Gothic	Royal	10	6	.012	.112	.087		Sans Serif
[1]Modern Gothic	Remington Rand	9-1/3 or 10	5.4	.013	.130	.089		
[1]Pica 134 or 534	Remington Rand	10	6	.010	.117	.074		
[1]Boldface #16*	IBM	32 units/inch	5-1/4	.013	.114	.114	.108	Executive Type (Prop. Spacing)
[1]Modern #32	IBM	32 units/inch	5-1/4	.015	.100	.113	.091	Executive Type (Prop. Spacing)
[1]Copperplate Gothic #46	IBM	36 units/inch		.014	.100	.110	.092	Executive Type (Prop. Spacing)
Artisan 10*	IBM	10	6	.014	.117	.090	.104	
Manifold 10	IBM	10	6	.014	.117	.090	.104	
Prestige Pica	IBM	10	5-1/4	.012	.110	.081	.098	
Accounting Gothic	IBM	10	6	.015	.106	.091	.0905	
Financial Gothic*	Underwood	10	6	.010	.100	.080		Farrington MX2021( )/G vocabulary
Modified Manifold 12*	IBM	12	6	.012	.100	.068	.090	Link Page Reader vocabulary
X3.1 (Size X)*	Underwood	10	5	.015	.125	.072		ASA Candidate Font for optical scanning

[1]These fonts are listed in Federal Supply Catalog under Air Force Stock Class 7340.

*Indicates carbon one-time ribbon used to prepare all samples.

TABLE I. CANDIDATE FONTS

Boldface #16 consists of upper case, lower case, and numerals, and has a pleasing appearance partly due to proportional character spacing and generous line spacing. Line spacing is considered adequate for OCR when, regardless of which characters are adjacent in two successive lines, a horizontal white band of at least 0.010 inches exists between lines. Variations among typewriters, however, must be taken into account. Another characteristic of Boldface #16, but one which might indicate limited usage, is that it is found only on electric typewriters having proportional spacing and utilizing a carbon ribbon. This implies that all characters will consist of uniform black strokes and a minimum number of voids.

Electronic masks were made for every character in the unstylized fonts investigated, along with cross-reference charts indicating the number of existing differences when the electronic mask of any one character is compared with the total form of any other character in a particular group. It was found that the total number of differences is at least five in nearly all cases.

The second candidate selected for investigation is Artisan 10. This unstylized font also includes upper case alphas, lower case alphas, and numerals, and is considered superior to Boldface #16 for optical scanning application. The data derived from actual type samples using the Automatic Character Analyzer revealed greater uniformity of stroke width for Artisan 10. The Artisan 10 font also has favorable line spacing and character height (six lines per inch and 0.117 inch, respectively). This guarantees adequate separation between lines under normal circumstances.

The characters in the Artisan 10 font were separated into three groups according to their gross physical characteristics.

Each electronic mask in this font (considering each group individually) has a minimum difference count of five (with two exceptions) when compared with the total form of each and every other character. The exceptions are in Group I where lower case "i" and the capital "I" provide poor discrimination.

Poor Discrimination Combinations:	"i" vs. "I" "I" vs. "i"
Artisan 10(1) System Scanning Resolution:	25 rows × 10 columns = 480 bits total
Artisan 10(2) System Scanning Resolution:	30 rows × 16 columns = 480 bits total
Artisan 10(1) Hardware Evaluation:	None
Artisan 10(2) Hardware Evaluation:	

Vocabulary size	62 characters
Number of documents read	37
Total number of characters	51,809
Total number of errors	37 (0.071%)
Total number of rejects	11 (0.021%)
Machine chargeable errors	20 (0.038%)
Machine chargeable rejects	7 (0.031%)

Errors and rejects that were not considered machine chargeable included the following:

1) "i" substitution for colons and semi-colons (not in vocabulary).

2) Substitutions or rejects which occurred for the open parenthesis and close parenthesis symbols (not in vocabulary).

3) Substitutions or rejects which occurred when characters were underlined.

NOTE: Most documents were read twice to complete volume requirements.

3. SEMI-STYLIZED FONT INVESTIGATIONS

A semi-stylized type font is one in which machine readability influenced its original design to a small degree or is a "conventional" font which has been modified in an attempt to better satisfy machine readability considerations. Most, if not all, fonts in this category have been created by the latter process — that of modification.

The choice of existing semi-stylized fonts is not extensive. One, a modified version of Underwood Financial Gothic (also referred to as Farrington Selfchek 12F1 numerals plus Farrington Selfchek 12H1 upper and lower case alphas and punctuation) constitutes the vocabulary of one of the early operational page readers, the Farrington MX2021( )/G. A second, Modified Manifold 12, is part of the vocabulary of a Link Page Reader used as a test vehicle in the subject program. Both of these fonts are characterized by good horizontal and vertical character separation. Also, shapes of certain characters in the unstylized versions of each font have been altered slightly in cases where area discrimination was poor when compared to certain other characters. Underwood Financial Gothic, however, is considered less suitable for an area analysis recognition technique because of narrow stroke width on nearly all available type samples. Although specified to be 0.012 inches ±0.003 inches, very few samples were found to have stroke widths as much as 0.012 inches. Very narrow strokes are considered undesirable, using an area analysis technique, because registration tolerance is reduced on incremental areas normally defined as being located well within the stroke boundaries.

To compensate by reducing the size of these incremental areas (i.e., finer scanning resolution) can only be done at the expense of noise tolerance. That is, the capability for ignoring small voids is reduced when scanning resolution is increased.

Electronic masks were made for every character in the semi-stylized fonts investigated along with cross-reference charts indicating the number of differences between encoded and total character forms.

The cross-reference charts for Underwood Financial Gothic show that, with the exception of four cases, every encoded character has at least five differences between each and every character combination in the respective groups. Group I, determined by the classification filter, consists of lower case "i" and lower case "l". Only one reliable difference was found to exist between these two characters. In Group II, the numeral "1" vs. the capital "I" and vice-versa show only three differences. Of the two semi-stylized candidate fonts then, Modified Manifold 12 is considered superior and was selected for hardware evaluation.

The characters in this font were separated into two groups according to their gross physical characteristics.

Each electronic mask in this font (considering each group individually) has a minimum difference count of five when compared with the total form of each and every other character.

Poor Discrimination Combinations:	None
System Scanning Resolution:	30 rows × 16 columns = 480 bits total

Hardware Evaluation:

Vocabulary Size	55 characters
Number of documents read	50
Total number of characters	59,052
Total number of errors	39 (0.069%)
Total number of rejects	27 (0.045%)
Machine chargeable errors	39 (0.069%)
Machine chargeable rejects	27 (0.045%)

4. STYLIZED FONT INVESTIGATIONS

A type font that is specifically designed to increase the reliability of an optical character recognition system with only secondary regard to aesthetics can be termed stylized. Characters in a stylized font are usually designed with bold and uniform strokes offering increased area differences between character combinations.

Some unstylized fonts have only a minimum area difference between certain characters, for example "B" and "8". Small area difference allows reliable recognition only under conditions of good print quality. Degradation along the leftmost vertical stroke of the "B" might cause significant dissimilarities to disappear. Excessive ink blooming in the left, center region of the numeral 8 will also increase the probability of substitution. Another example is the upper case "S" versus the numeral "5". Also, in some fonts of the unstylized or even semi-stylized class, insufficient area differences exist between the upper case "I" and the numeral "1".

Subcommittee X3.1 on Character Recognition under American Standards Association Sectional Committee X3 has designed a stylized font, intended for optical character recognition application. Known as the TG1C font to that organization, this font includes upper case alphas and numerals and offers improved area difference over unstylized counterparts. Cases where insufficient area differences often exist have been given special consideration. Distinct differences were provided between "B" and "8", "I" and "1", "5" and "S", and "Oh" and zero such that a greater degree of mutilation can be tolerated without misidentification by the reading equipment. Stroke widths are uniform and separation between any combination of characters in this font is adequate, providing typewriter keys are not bent.

The ASA Candidate Font is a single case font, with proposed availability in four basic sizes, designated W, X, Y, and Z, where W is the smallest and Z the largest.

In the interest of promoting font standardization, the Rome Air Development Center has purchased a typewriter with size X. Samples from this machine provide the basis for the stylized font investigation described.

The characters in this font were regarded as one group with the exception of the capital "I". The "I" being narrower than any other character in the vocabulary permitted easier recognition by width and height measurement.

Each electronic mask proved to have a minimum difference count of five when compared with the total form of each and every other character.

Poor Discrimination Combinations: None

System Scanning Resolution: 30 rows × 16 columns = 480 bits total

Hardware Evaluation:

Vocabulary size	36 characters
Number of documents read	33
Total number of characters	54,749
Total number of errors	39 (0.071%)
Total number of rejects	0 (0%)
Machine chargeable errors	2 (0.004%)
Machine chargeable rejects	0 (0%)

Errors and rejects that were not considered machine chargeable included the following:

1) Overlapped characters caused by typist error.
2) Characters not considered part of the ASA Candidate Font.
3) Symbols not included in the ASA Candidate Font.

NOTE: Most documents were read three times to complete volume requirements.

SECTION V

RELATED FEATURES

1. REFLECTANCE TESTS

Significant improvements in the "quality" (clearness, sharpness, cleanness, etc.) of typewritten character impressions can be obtained by careful selection of both paper and ribbon. For optimum reading performance, documents should also be characterized by high contrast between character impressions and paper. Careful selection of paper and ribbons helps to ensure adequate contrast and is especially important when reusable ribbons are employed in the machines producing the documents to be scanned. Reflectance measurements within the range of spectral sensitivity of the reader photosensors provide, therefore, one important basis for paper and ribbon selection.

a. Reflectance Measuring Equipment

The response of the photosensors (T. I. Type LS400) used in the Link Page Reader peaks sharply at 900 millimicrons; therefore, the reflectance measuring apparatus shown in Figure 12 is designed to be effective only in the vicinity of this wavelength. Most of the system is composed from commercially available components. Reflectance measurements have been found to agree closely with costly recording spectrophotometers.

A cutaway view (Figure 13) shows the complete equipment consisting of four units:

1) The measuring sphere, including photomultiplier.
2) The indicating instrument, which is used as a microammeter.
3) The power supply and amplifier.
4) The tungsten light source with filters.

Paper samples are inserted beneath the cover on top of the integrating sphere using a nonreflective black backing material. Samples are compared against a "white" working standard having a uniform stable reflectivity within the range of 400 to 1,100 millimicrons. The standard of total reflectance is magnesium carbonate, which is taken as 100 percent. A magnesium carbonate block, however, is inconvenient for calibration and comparison. Therefore, the device employs a white porcelain working standard. This porcelain has less reflectance than magnesium carbonate and should not be used for setting of 100 percent reflectance when taking measurements. If, however, the light control is set for true reflectance of the porcelain at a given light wavelength and the porcelain is replaced with the sample, the reading will be the correct reflectance of the sample. A block of magnesium carbonate is used for calibrating the porcelain standard periodically. The equipment calibration procedure is to:

1) Set the wavelength of interest by selecting a filter. Presently used is a Corning Filter CS7-69 with a cutoff wavelength of 0.85 u.

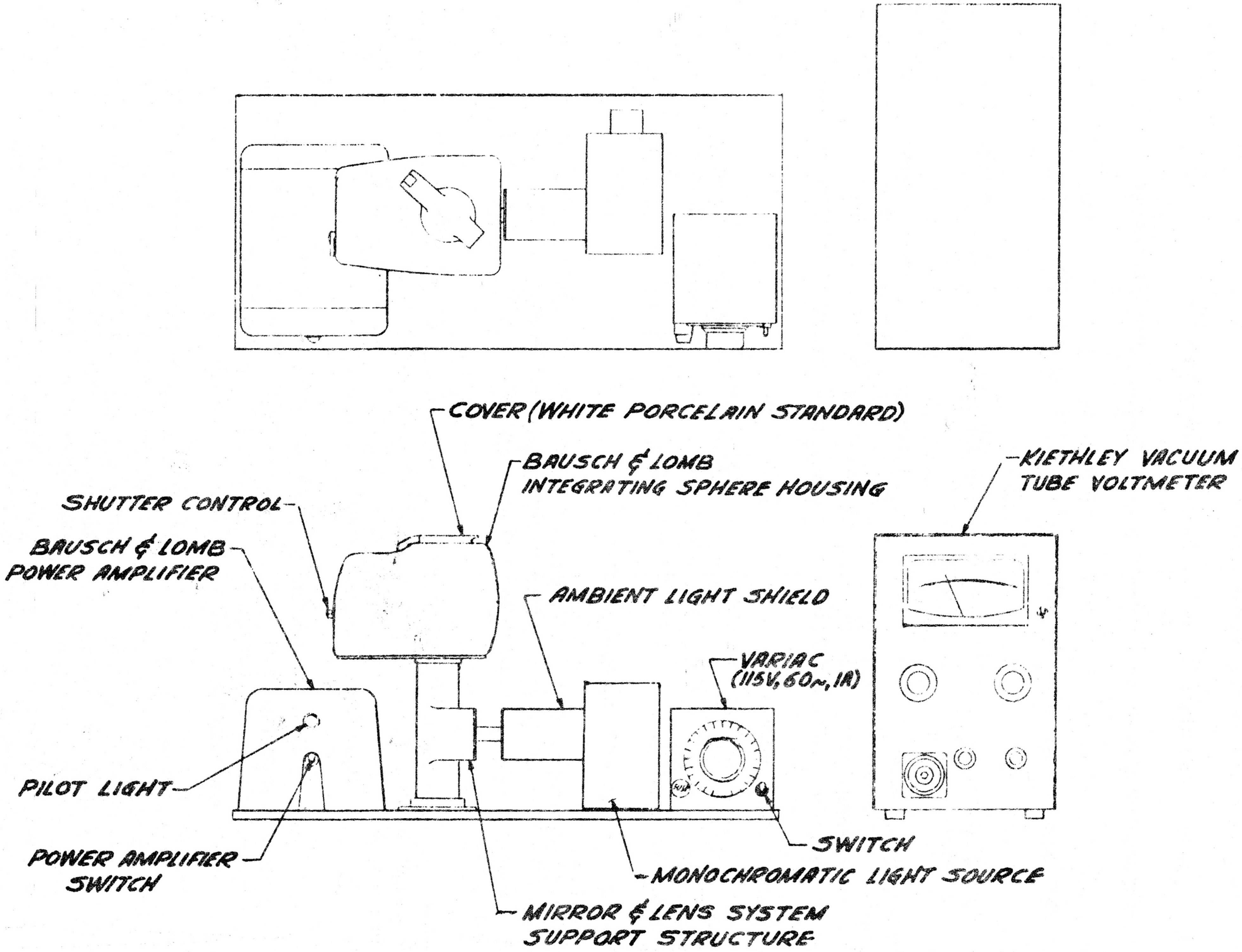

Figure 12. Link Reflectance Measuring Equipment

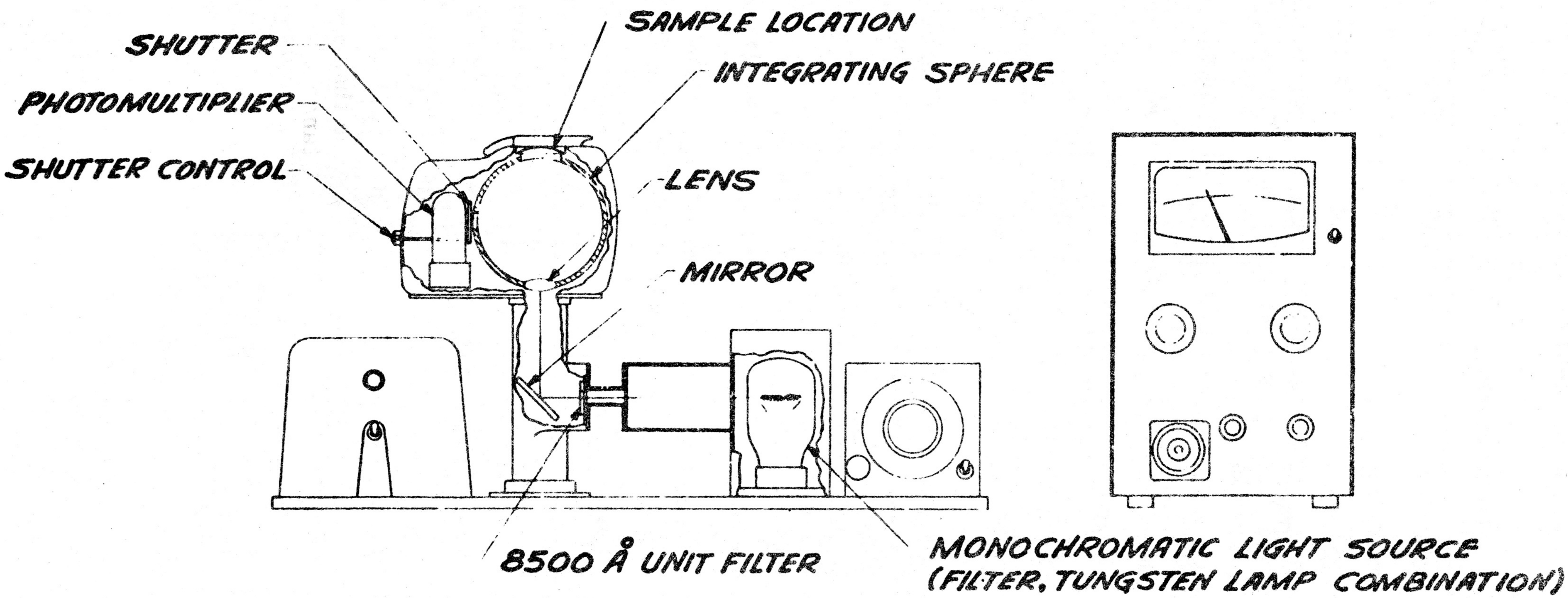

Figure 13. Internal Configuration of Link Reflectance Measuring Equipment

2) Adjust microammeter for zero percent reflectance.

3) Place the magnesium carbonate block on the measuring aperture.

4) Open the light shutter and adjust the lamp control (Variac) for 100 percent reflectance.

5) Close the light shutter and substitute the porcelain standard for the magnesium carbonate.

6) Open the light shutter and read the reflectance of the porcelain. Record this reading and check it before each individual measurement.

By following this procedure for each wavelength of interest, a reflectance curve of porcelain secondary standard can then be made at any wavelength. Because the block of magnesium carbonate is satisfactory as a standard only as long as it is clean and not aged, it should be scraped lightly with a knife if it becomes stained.

b. Paper Reflectance Measurements

The Reflectance Measuring Equipment has been employed to determine reflectances of a large assortment of paper samples. Table II lists the percent reflectance of each sample tested and identifies the paper as to manufacturer and weight.

c. Ribbon Impression Reflectance Measurements

The measurement of rough texture samples such as typewriter ribbons can also be accomplished with the Link Reflectance Measuring Equipment because the measuring area is large enough (about 0.5 square inch) to provide test results of high accuracy and reproducibility.

The results of a comparison test of five Columbia ribbons are given in Table III. The reflectance measurements for each ribbon were taken on samples prepared by typing an identical dense pattern of a special character (i.e., an Item Separator Symbol in the Modified Manifold 12 type font) on sheets of paper of various grades.

The patterns were kept as nearly identical as possible by using a single electric typewriter with the same impression and multiple copy control settings. Further control was exercised by taking the average of the readings from three separate samples of each paper-ribbon combination.

These results (Table III) give an accurate indication of the range of reflectivity which can be expected with the ribbons involved. However, several factors such as voiding, smearing, and other ink transfer characteristics may influence the reading without giving an absolute indication of the contribution of each. For this reason, reflectivity alone is not a fool-proof indicator of the suitability of a given ribbon for optical scanning, although it does quickly identify those which are unsuitable from the viewpoint of insufficient contrast at the particular wavelengths of interest. Thus, reflectivity is only one of several characteristics of interest in the selection of suitable ribbons for optical scanning as discussed in Section V. 1.d.

TABLE II. PAPER REFLECTANCE MEASUREMENTS

Sample No.	Manufacturer	Wgt.	Description	Reflectance (percent)
1	Columbia	16	Trojan Bond	68.5
2	Columbia	20	NSI-418 S/C	81.3
3	Columbia	24	TSI-1207	84.7
4	Columbia	24	WC-2-8008A	86.5
5	Columbia	24	Mayville	82.0
6	Mead	11	Opaque Form Bond	65.5
7	Mead	20	Moist-Rite Bond	79.0
8	Mead	24	Moist-Rite Bond	81.3
9	Mead	33	Opaque Circular	77.4
10	Mead	40	Opaque Circular	80.9
11	Oxford	18	MC-2-3602	87.5
12	Oxford	19	Rangely PM2-3035	84.0
13	Oxford	20	US2-2578	83.3
14	Oxford	20	WC-2-4865A	83.0
15	Oxford	20	MC2-5114	82.5
16	Oxford	20	SCAN-51-YY1	84.7
17	Oxford	20	PM2-3997	85.1
18	Oxford	20	US2-2476	81.8
19	Oxford	20	X-3539H S/C	78.8
20	Oxford	28	WC-2-7018A	89.8
21	Howard	16	Vellum Bond	83.9
22	Miami Paper Co.	16	Coated Bond	86.4
23	Finch, Pruyn & Co.	20	Finch Offset	77.2
24	Aetna Paper Co.	20	English Maxopaque	80.8
25	Standard Paper Co.	20	Surgave Plate	74.1
26	Federal Spec. UU-P-121-1 Type III	20	Unknown	80.7
27	Mead	20	Duplicator	77.4

TABLE III. RIBBON IMPRESSION REFLECTANCE MEASUREMENTS

Sample No.	Paper Manufacturer	Wgt.	Ribbon	Pattern Reflectance (percent)
1	Columbia NSI-418 S/C	20	Columbia M50LX082	12.0
2	Oxford MC 2-3602	18	Columbia M50LX082	14.3
3	Oxford PM 2-3035	18	Columbia M50LX082	15.3
4	Oxford US 2-2578	20	Columbia M50LX082	14.0
5	Oxford WC 2-8008A	24	Columbia M50LX082	15.0
6	Oxford WC 2-4865A	20	Columbia M50LX082	13.7
7	Columbia NSI-418 S/C	20	Columbia SF-50	25.7
8	Oxford MC 2-3602	18	Columbia SF-50	29.0
9	Oxford PM 2-3035	18	Columbia SF-50	29.7
10	Oxford US 2-2578	20	Columbia SF-50	23.3
11	Oxford WC 2-8008A	24	Columbia SF-50	24.0
12	Oxford WC 2-4865A	20	Columbia SF-50	25.3
13	Columbia NSI-418 S/C	20	Columbia M50LX090	14.5
14	Oxford MC 2-3602	18	Columbia M50LX090	14.7
15	Oxford PM 2-3035	18	Columbia M50LX090	16.2
16	Oxford US 2-2578	20	Columbia M50LX090	16.0
17	Oxford WC 2-8008A	24	Columbia M50LX090	13.3
18	Oxford WC 2-4865A	20	Columbia M50LX090	13.3
19	Columbia NSI-418 S/C	20	Columbia PF75-P53	14.8
20	Oxford MC 2-3602	18	Columbia PF75-P53	18.3
21	Oxford PM 2-3035	18	Columbia PF75-P53	15.7
22	Oxford US 2-2578	20	Columbia PF75-P53	15.5
23	Oxford WC 2-8008A	24	Columbia PF75-P53	15.2
24	Oxford WC 2-4865A	20	Columbia PF75-P53	17.2
25	Columbia NSI-418 S/C	20	Columbia SF-730	21.0
26	Oxford MC 2-3602	18	Columbia SF-730	23.5
27	Oxford PM 2-3035	18	Columbia SF-730	22.2
28	Oxford US 2-2578	20	Columbia SF-730	21.3
29	Oxford WC 2-8008A	24	Columbia SF-730	22.8
30	Oxford WC 2-4865A	20	Columbia SF-730	25.7
31	Howard	16	IBM 5121	13.0

d. Evaluation of Carbon Ribbon and Paper Combinations

During the course of this study it became apparent that the best combination for optical scanning application was not necessarily the paper having the highest reflectance and opacity combined with the ribbon which produced the lowest reflectance character strokes, although these are, of course, important characteristics. For optimum results, properties such as voids, addition noise (ink splatter), and contrast should be evaluated using various paper and ribbon combinations.

A test covering all possible combinations of commercially available ribbons and paper would be a formidable, if not impossible, task. Therefore, only a limited number of possibilities were tried from which several acceptable combinations resulted. Only carbon ribbons were tested and these were limited to polyethylene and mylar types. Most of the paper considered was 20 pound (basis - 500 sheets, 17 inches by 22 inches) and all had reflectance exceeding 80 percent when measured using the Link Reflectance Measuring Equipment. All ribbons and paper used in the test are commercially available.

A single electric typewriter (IBM Model 11) was used to prepare all test samples using a medium setting (five) on the impression control and setting "A" on the multiple copy control. Each sample was made up using the same series of characters and symbols.

Tests were conducted with the paper samples divided into two major groups in anticipation of several ribbons being consistently deficient in one or more respects. Sample preparation and testing of deficient ribbons using the second group of paper samples was, therefore, considered unnecessary.

Test data for the first group of paper samples is given in Figures 14 through 17 with similar data given for the second group in Figures 18 through 21. In Figures 14 and 18 are recorded actual photodiode signal amplitudes (millivolts) resulting from scanning the test documents using the Model X-3 Scanner. This gives a very good indication of relative contrast for the various ribbon-paper combinations and from this it is easily seen that the SF-50, SF-100, and SF-730 ribbons produce output signals generally inferior to the other ribbons tested.

Figures 15 and 19 show the results of a visual examination of addition noise (ink splatter, fuzziness of stroke edges, smear, etc.). From this it can be seen that the M50 LX090 and SF-730 ribbons are generally unacceptable. Severe flaking of the SF-730 was evident on all samples even without examination under magnification.

Figures 16 and 20 give an indication of the number of voids which occur within the character strokes using the various combinations of ribbon and paper. The M50 LX090 and SF-730 ribbons were also generally unacceptable in this respect.

Figures 17 and 21 show stroke width measurements with the most interesting result being the significantly wider strokes produced by the SF-100 special mylar

Paper ↓ / Ribbon →	IBM 5121 (Polyethylene)	Columbia SF-50 (Mylar)	Columbia M50 LX090 (Mylar)	Columbia PF75 P53 (Polyethylene)	Columbia SF-100 (Special Mylar)	Columbia M50 LX082 (Mylar)	Columbia SF-730 (Mylar)
Oxford WC-2-4865A (20 LB - 83%)	750	650	800	800	600	700	600
Oxford WC-2-8008A (24 LB - 86.5%)	800	650	800	800	600	750	680
Oxford US2-2578 (20 LB - 83.25%)	800	600	800	800	600	750	650
Oxford PM2-3035 (18 LB - 84%)	750	600	750	750	600	750	680
Oxford MC2-3602 (18 LB - 87.5%)	750	550	750	700	550	750	650
Columbia Scanamaster NSI-418 S/C (20 LB - 81.25%)	800	600	750	800	600	750	

Test: Relative Contrast (millivolts photodiode signal)
Light Level: 70 volts
Photodiode Reference: #24

Figure 14. Group 1 Paper-Ribbon Evaluation (Relative Contrast)

Paper ↓ / Ribbon →	IBM 5121 (Polyethylene)	Columbia SF-50 (Mylar)	Columbia M50 LX090 (Mylar)	Columbia PF75 P53 (Polyethylene)	Columbia SF-100 (Special Mylar)	Columbia M50 LX082 (Mylar)	Columbia SF-730 (Mylar)
Oxford WC-2-4865A (20 LB - 83%)	Little	Little	Severe	Little	Moderate	Moderate	Severe
Oxford WC-2-8008A (24 LB - 86.5%)	Little	Little	Severe	Moderate	Moderate	Moderate	Severe
Oxford US2-2578 (20 LB - 83.25%)	Moderate	Little	Moderate	Moderate	Moderate	Moderate	Severe
Oxford PM2-3035 (18 LB - 84%)	Little	Little	Moderate	Very Little	Little	Moderate	Severe
Oxford MC2-3602 (18 LB - 87.5%)	Moderate	Little	Severe	Severe	Moderate	Little	Moderate
Columbia Scana-master NSI-418 S/C (20 LB - 81.25%)	Little	Little	Severe	Very Little	Little	Very Little	Moderate

Test: Addition Noise (after careful handling)
Light Level:
Photodiode Reference:

Figure 15. Group 1 Paper-Ribbon Evaluation (Addition Noise)

Paper ↓ / Ribbon →	IBM 5121 (Polyethylene)	Columbia SF-50 (Mylar)	Columbia M50 LX090 (Mylar)	Columbia PF75 P53 (Polyethylene)	Columbia SF-100 (Special Mylar)	Columbia M50 LX082 (Mylar)	Columbia SF-730 (Mylar)
Oxford WC-2-4865A (20 LB - 83%)	Many	Few	Many	Many	Few	Many	Very Many
Oxford WC-2-8008A (24 LB - 86.5%)	Many	Few	Many	Few	Few	Few	Very Many
Oxford US2-2578 (20 LB - 83.25%)	Many	Few	Many	Few	Very Few	Few	Very Many
Oxford PM2-3035 (18 LB - 84%)	Many	Few	Many	Many	Very Few	Few	Very Many
Oxford MC2- 3602 (18 LB - 87.5%)	Many	Many	Many	Many	Very Few	Many	Very Many
Columbia Scana-master NSI-418 S/C (20 LB - 81.25%)	Few	Many	Many	Many	Few	Many	Very Many

Test: Voids
Light Level:
Photodiode Reference:

Figure 16. Group 1 Paper-Ribbon Evaluation (Voids)

Ribbon → / Paper ↓	IBM 5121 (Polyethylene)	Columbia SF-50 (Mylar)	Columbia M50 LX090 (Mylar)	Columbia PF75 P53 (Polyethylene)	Columbia SF-100 (Special Mylar)	Columbia M50 LX082 (Mylar)	Columbia SF-730 (Mylar)
Oxford WC-2-4865A (20 LB - 83%)	.013	.014	.014	.014	.016	.013	.014
Oxford WC-2-8008A (24 LB - 86.5%)	.013	.014	.014	.014	.016	.013	.014
Oxford US2-2578 (20 LB - 83.25%)	.013	.015	.014	.014	.016	.013	.015
Oxford PM2-3035 (18 LB - 84%)	.013	.015	.014	.014	.016	.013	.014
Oxford MC2-3602 (18 LB - 87.5%)	.013	.014	.014	.014	.015	.013	.015
Columbia Scana-master NSI-418 S/C (20 LB - 81.25%)	.012	.014	.013	.013	.015	.013	.014

Test: Stroke Width
Light Level:
Photodiode Reference:

Paper ↓ / Ribbon →	IBM 5121 (Polyethylene)	Columbia PF 75 P53 (Polyethylene)	Columbia M50 LX082 (Mylar)
Oxford US2-2476 Grade OE - Grain Short (20 LB - 81.8%)	800	750	750
Oxford MC2-5114 Grade MWLD - Grain Short (20 LB - 82.5%)	800	800	850
Oxford Scan-51-YY1 #20 Scanamaster (20 LB - 84.7%)	800	850	850
Oxford PM2-3997 - Grade RT - Grain Long (20 LB - 85.1%)	800	750	800

Figure 18. Group 2 Paper-Ribbon Evaluation (Relative Contrast)

Test: Relative Contrast (millivolts photodiode signal)
Light Level: 67 volts
Photodiode Reference: #24

Paper ↓ \ Ribbon →	IBM 5121 (Polyethylene)	Columbia PF75 P53 (Polyethylene)	Columbia M50 LX082 (Mylar)
Oxford US2-2476 Grade OE - Grain Short (20 LB - 81.8%)	Little	Little	Moderate
Oxford MC2-5114 Grade MWLD - Grain Short (20 LB - 82.5%)	Little	Little	Little
Oxford Scan-51-YY1 #20 Scanamaster (20 LB - 84.7%)	Little	Little	Moderate
Oxford PM2-3997 Grade RT - Grain Long (20 LB - 85.1%)	Little	Little	Little

Figure 19. Group 2 Paper-Ribbon Evaluation (Addition Noise)

Test: Addition Noise (after careful handling)
Light Level:
Photodiode Reference:

Paper ↓ / Ribbon →	IBM 5121 (Polyethylene)	Columbia PF 75 P53 (Polyethylene)	Columbia M50 LX082 (Mylar)
Oxford US2-2476 Grade OE - Grain Short (20 LB - 81.8%)	Many	Many	Many Severe
Oxford MC2-5114 Grade MWLD - Grain Short (20 LB - 82.5%)	Few	Few	Many Severe
Oxford Scan-51-YY1 #20 Scanamaster (20 LB - 84.7%)	Few	Very Few	Many Severe
Oxford PM2-3997 Grade RT - Grain Long (20 LB - 85.1%)	Many	Many	Many Severe

Figure 20. Group 2 Paper-Ribbon Evaluation (Voids)

Test: Voids
Light Level:
Photodiode Reference:

Paper ↓ / Ribbon →	IBM 5121 (Polyethylene)	Columbia PF 75 P53 (Polyethylene)	Columbia M50 LX082 (Mylar)
Oxford US2-2476 Grade OE - Grain Short (20 LB - 81.8%)	.014	.014	.014
Oxford MC2-5114 Grade MWLD - Grain Short (20 LB - 82.5%)	.012	.013	.013
Oxford Scan-51-YY1 #20 Scanamaster (20 LB - 84.7%)	.013	.013	.013
Oxford PM2-3997 Grade RT - Grain Long (20 LB - 85.1%)	.013	.013	.014

Figure 21. Group 2 Paper-Ribbon Evaluation (Stroke Width)

Test: Stroke Width
Light Level:
Photodiode Reference:

ribbon. An optical scanner designed to read documents produced using this ribbon should take this feature into account.

Careful examination of all data indicates that no more than five of the 42 ribbon and paper combinations are acceptable in all respects for optical scanning purposes. Listed in order of preference, these are:

Ribbon	Paper
Columbia PF75 P53	Oxford Scan-51-YY1 #20 Scanamaster
Columbia PF75 P53	Oxford MC2-5114 Grade MWLD
IBM 5121	Oxford Scan-51-YY1 #20 Scanamaster
IBM 5121	Oxford MC2-5114 Grade MWLD
IBM 5121	Columbia Scanamaster NSI-418 S/C

The differences among the above combinations are extremely small and the order of preference might easily be altered by other observers. The Columbia PF75 P53 ribbon exhibited slightly superior resistance to smear when the test documents were subjected to much manual handling; therefore, it is preferable to the IBM 5121 where this characteristic is important. The Columbia Scanamaster NSI-418 S/C is reportedly a calendered version of Oxford Scan-51-YY1 #20 but otherwise identical. Calendering seems to give no apparent improvement and only results in a more transparent paper resulting in lower reflectance measurements and slightly reduced contrast. The Scan-51-YY1, MC2-5114 and Scanamaster NSI-418 S/C samples were the only coated papers tested for ribbon compatibility. All other papers were uncoated.

e. Preprinted Form Inking

Preprinted text or lines on the documents which have no significance to the Scanner should ideally have the same reflectivity as the paper itself; they will then be "invisible" to the Scanner. It has been found that a correct mixture of red and white inks can simultaneously be highly visible to the human eye and practically invisible to the Scanner.

Two excellent "dropout" colors are Splended Red, mixed one to ten with white, and ML 117, made by Van Son Holland, Inc., Mineola, New York. A reflectivity diagram of the latter is shown in Figure 22.

For maximum efficiency, forms should be arranged so that the information to be read is confined to consecutive lines to facilitate machine programming. This also minimizes the time lost in searching for data or skipping fields.

2. CHARACTER IMPRESSION TOLERANCES

For an area analysis recognition technique to function effectively, characters to be read must be characterized by certain standards of angular orientation, registration, dimensional stability, additive noise, reductive noise, and contrast.

In any data processing system employing OCR, there are those who are concerned as to whether these standards are met. These include the printing device manufacturer,

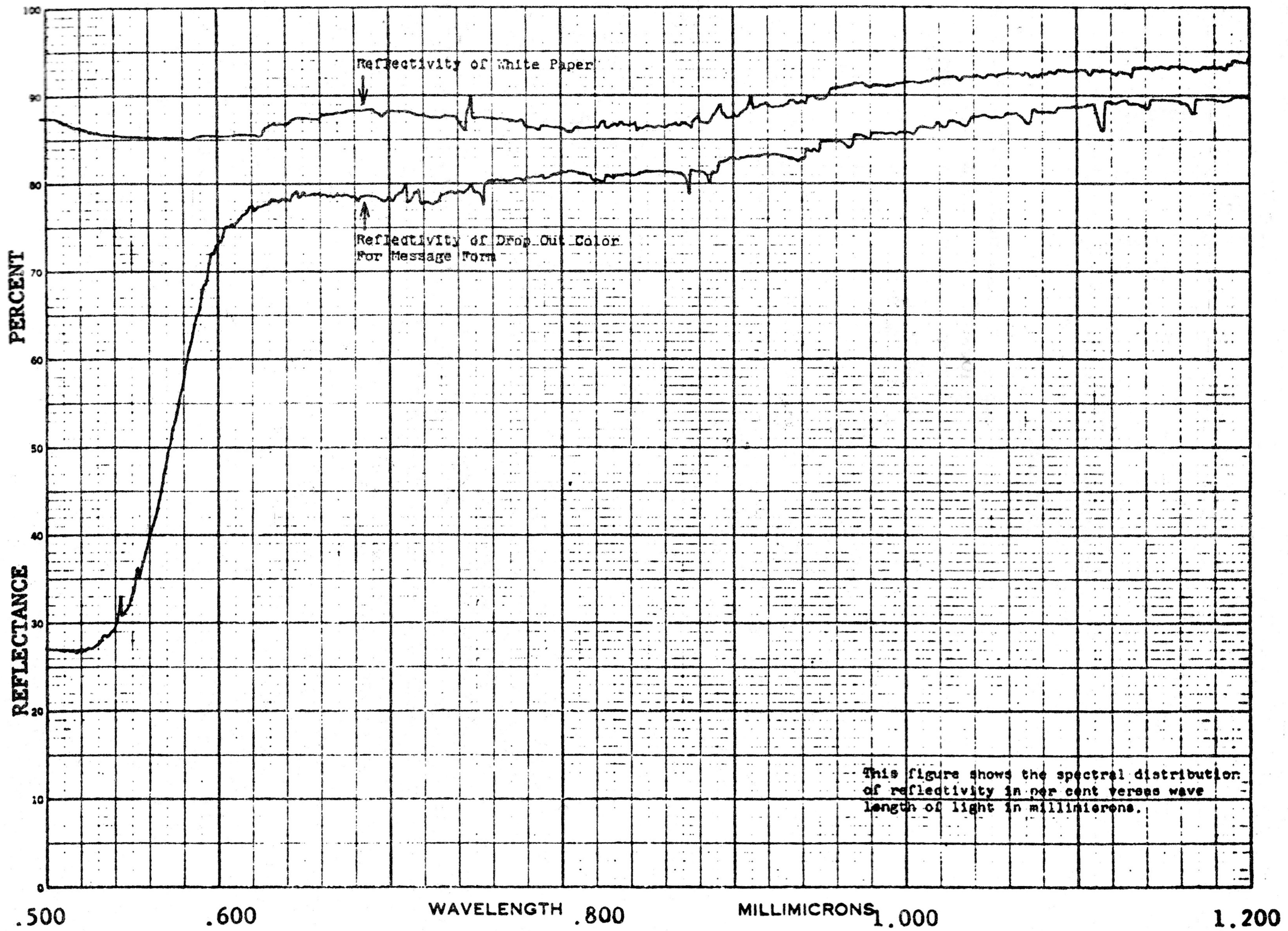

Figure 22. Reflectivity Diagram

various printing establishments (when preprinted forms are involved), paper manufacturers, OCR machine manufacturer, and OCR machine user. Each is naturally concerned as to whether his product is compatible with OCR character quality requirements.

For OCR equipment to accommodate a majority of documents prepared using a variety of printing devices including typewriters, posting/bookkeeping machines, and high-speed computer printer, the equipment should be characterized by the following:

Angular Orientation: Variations in character skew must be expected when scanning machine-prepared documents because of manufacturing tolerances, bent keys, and document skew during preparation.

In earlier vocabulary developments, it was determined that skew tolerance, while reading, nearly always extended at least 1.5 degrees beyond the range of skew included in the original typed impressions upon which the memory is based. Therefore, if the original impressions cover the range -1.5 degrees to +1.5 degrees, a total skew tolerance of ±3 degrees must result. The -1.5 to +1.5 degree range on the original impressions was induced artifically by clockwise and counterclockwise rotation of the image transducer. Counterclockwise rotation of the transducer simulates clockwise rotation of the characters and conversely, clockwise rotation of the transducer simulates counterclockwise character rotation.

To increase skew tolerance, all character encoding has been based on data recorded at +1.5 degree and at -1.5degree artificially-induced skew, the objective being a final skew tolerance in the reading machine of at least ±3 degrees.

Sufficient invariant video data was available in all cases which resulted in just one encoded form of each character (a double coding becomes necessary in many cases if the skew tolerance is required to exceed ±3 degrees).

Experimental reading tests have verified that a final ±3 degree skew tolerance was achieved. This is the recommended maximum tolerance without placing undue requirements on OCR equipment.

Registration: OCR equipment should readily accommodate intermixed, both fixed and proportional, horizontal character spacing up to a maximum of at least 12 characters per lineal inch and vertical line spacings which range between five and six lines per inch. Separation between adjacent characters as small as 0.010 inches should be permissible. Misregistration of adjacent characters should be tolerated up to plus or minus one-half the height of a nominal size character.

Dimensional Stability: For area analysis techniques, extreme stroke width variations as well as character height variations should be avoided. This is because most systems of this type rely on incremental white and black areas being consistently located for correct identification. The variation in stroke width between the minimum and the maximum allowable width creates areas of uncertainty, which cannot be used as a basis for recognition. In the hardware evaluation of fonts, character impressions upon which the reading machine memory is based were imprinted with low- and high-impression settings, then superimposed to provide a definition of typical limits on such areas. Based on the results of this evaluation, it has been established that

stroke width should be maintained to within ±0.003 inch of nominal. The suggested nominal value is 0.013 inches for fonts in which character heights fall within the range 0.100 to 0.130 inches.

Noise: Voids within the strokes of any character of the Scanner vocabulary should not exceed the area of a square 0.004 inches per side. Also, the concentration of permissible size voids should not be great enough to increase the diffuse reflectance of any minimum width character stroke above 20 percent between 400 and 1,100 millimicrons referred to magnesium carbonate as the primary white standard and measured against a flat black background whose reflectance does not exceed four percent (3M Flat Black coating).

Dirt or other extraneous markings on the documents should not exceed the area of a square 0.004 inches per side when located within 0.010 inches of a rectangle which just encloses any character to be read (this is the nominal size of the scanning field). In other areas, extraneous marks should be limited to the area of a square 0.010 inches per side. The concentration of extraneous dirt or other markings should not be great enough to reduce the diffuse reflectance of any square background area 0.100 inches per side below 80 percent between 400 and 1,100 millimicrons referred to magnesium carbonate as the primary white standard and measured against a flat black background whose reflectance does not exceed four percent (3M Flat Black coating).

Contrast: Contrast variation on OCR documents is usually a result of reduced character blackness with ribbon usage and/or use of paper with varying reflectance characteristics. This problem has been minimized in evaluating reading performance in the subject program since all documents were prepared on one grade of paper using electric typewriters equipped with single-use carbon ribbon.

Previous investigations, however, have shown efforts made toward increasing and preserving contrast are well worthwhile since a resultant improvement in signal-to-noise ratio in the OCR equipment is made possible. With OCR equipment which operates in the range 400 to 1,100 millimicrons, good contrast can be ensured by employing paper, ribbons, and inks as recommended in Sections V. 1.d. and V. 1.e.

SECTION VI

CONCLUSIONS

On the basis of these tests, the following can be concluded:

1. It has been established that a stylized font offers the greatest advantages in applications where accuracy of the OCR equipment is of paramount importance.

2. In general, OCR equipment designed to read a single case font will provide superior performance over equipment designed to read a double case vocabulary.

3. Using unmutilated documents having good character quality, combined error/reject rates significantly lower than 0.2 percent have been demonstrated for unstylized, semi-stylized, and stylized fonts.

4. With OCR equipment which operates in the range 400 to 1,100 millimicrons, the minimum diffuse reflectance of the paper should be 80 percent or greater referred to magnesium carbonate as the primary white standard and measured against a flat black background. A variety of commercially available papers will meet this specification.

5. Reflectance of typed impressions on such paper should not exceed 20 percent in order to provide sufficient contrast at the wavelengths of interest.

6. The acceptability of particular paper, ink, or ribbon should not be judged by reflectivity alone; i.e., combinational effects should also be considered.

7. To ensure the nonexistence of reading problems, character impressions in an OCR environment should have a minimum separation of 0.010 inches and be free from overhang or underhang.

8. For an area analysis reading technique, character skew (angular deviation from an erect position) should not exceed ±3 degrees, including the effects of line skew.

9. Line skew, which is mainly a function of typist proficiency, should not exceed one degree in order to avoid undue complexity in the OCR equipment.

10. To satisfy performance and appearance considerations, stroke width in an OCR font should be maintained within ±0.003 inches, or nominal. The recommended nominal width is 0.013 inch.

11. Height of character impressions should fall within the range of 0.100 to 0.130 inches.

12. Retyping after making <u>clean</u> erasures has been found to be an acceptable correction method for OCR. The use of an erasing shield is recommended to avoid disfiguration of adjacent characters.

SECTION VII

RECOMMENDATIONS

1. A technique for measuring print quality should be developed that is practical, inexpensive, and fast.

2. More exacting parameters should be determined for specifying papers and ribbons that are acceptable for optical character recognition use.

3. In each of the general categories of unstylized, semi-stylized, and stylized fonts, a recommended font has been presented, its selection justified, and its performance demonstrated using an existing optical page reader. On the basis of the fonts tested, these recommendations are summarized as follows:

FONT SELECTION GUIDE

Order of Preference	Font Classification	Characters	Recommended Font
1	Stylized	Single case	ASA Candidate
2	Semi-Stylized	Single case	Modified Manifold 12
3	Unstylized	Double case	Artisan 10

4. This investigation, although limited in scope because of low funding, shows the distinct advantages of using a stylized, machine-designed font, such as the American Standards Association Font. Comprehensive tests should now be funded which will test the latest version of the A. S. A. Font in a fully-developed and operational multifont reader.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Optical Character Recognition Fonts Standardization Assessment of Typewriter Fonts Optical Scanning						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through
_______________________________."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through
_______________________________."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through
_______________________________."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional

UNCLASSIFIED

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) Rome Air Development Center Griffiss Air Force Base, N. Y. 13440	2a. REPORT SECURITY CLASSIFICATION Unclassified 2b. GROUP

3. REPORT TITLE

Standardization of Typewriter Fonts for Automatic Reading

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report - February 1963 to September 1965

5. AUTHOR(S) (Last name, first name, initial)

Greenly, James F.

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
March 1966	53	0

8a. CONTRACT OR GRANT NO.	9a. ORIGINATOR'S REPORT NUMBER(S)
b. PROJECT NO. DW-63-48	RADC-TR-65-523
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)
d.	

10. AVAILABILITY/LIMITATION NOTICES

Distribution of this document is unlimited.

11. SUPPLEMENTARY NOTES	12. SPONSORING MILITARY ACTIVITY
	RADC

13. ABSTRACT

The objective of this program was to provide engineering data in support of standardization of typewriter fonts and related features for optical scanning application. Primary emphasis was placed on investigation and evaluation of existing typewriter fonts and includes an evaluation of a type font developed by Subcommittee X3.1 on Character Recognition under American Standards Association Sectional Committee X3. Investigations were by computer programmed assessment of each font using a technique developed partly under Contract AF30(602)-2642 sponsored by Rome Air Development Center and partly under continuing Link sponsored character recognition efforts. Evaluations were accomplished by extending the vocabulary capacity of a Link Multifont Page Reader to permit machine reading of a significant volume of typewriter-prepared documents. Reject and error rates were determined in this manner for each of several type styles considered.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification